Data-Aware Control $\Rightarrow$ Efficient Traffic Management

(Extended Abstract)

Goce Trajcevski Alok Choudhary
Department of Electrical Engineering and Computer Science
Northwestern University
Evanston, Il 60208
{goce,choudhar}@eecs.northwestern.edu

"We need to emphasize smarter roads – sensors that detect 'nonrecurring' traffic disruptions (the cause of an estimated one-third of traffic delays), for instance, and intelligent traffic signals and speed limits that react to changing conditions".

Tom Vanderbilt, author "Traffic" (*BusinessWeek, February 2, 2009*).

1. Introduction and Motivation

Almost every driver has lived through the frustrating experience of traffic congestions. While most of us have "learned to live with it" and avoid more or less predictable problems such as rush hour, parades, and big sports or political events, we are regularly caught by less predictable events and by events with unpredictable severity. One can easily list other similar events, such as severe weather conditions, sudden disruptions of the public transportation system or air-traffic delays causing unexpected disruptions of regular traffic patterns around airports.

The importance of the *Automatic Traffic Management* (ATM) has been recognized since the early 1900's – for instance, Ghiglieri's automatic traffic signals with red and green lights were introduced in California in 1917, and the first computer-based synchronized traffic lights in the US emerged in the 1970's. In the 1990's, the Federal Highway Administration (FHWA) established the Adaptive Control Software (ACS) research program [1, 2]. More recently, in 2004 the US Department of Transportation launched the VII (Vehicle Infrastructure Integration) initiative [3] with a vision of deploying a communications infrastructure on the roadways and in all production vehicles. Nevertheless, in the US alone, traffic congestion is responsible for billions of wasted man-hours (an average driver in Chicagoland will annually spend 56 hours in traffic delays); hundreds of billions of additional fuel being consumed annually (e.g., according to the Bureau of Transportation Statistics, 141.6 million gallons of fuel during 2007 only in Chicago [4]; and a severe environmental impact from pollution [2, 5]. Tragically, traffic accidents resulted 31,140 fatalities in the period of January-October 2008 alone [6], in addition to over 3 million personal injuries, and over $200 billion in property damages.

Methodologies, tools and systems aiming at efficient traffic management abound, however, the existing approaches rely on models based on past observations, *augmented* with traffic-related information (e.g., traffic flow) in *limited spatial regions*. Consequently, they are not able to fully exploit the knowledge of different *spatio-temporal correlations* among traffic data patterns, thereby limiting the spectrum of *control algorithms* that can be used in different contexts. In addition, the vast amount of real-time data (e.g., from road-sensors) becoming available in various installations is *not efficiently matched* with the historic data, thereby limiting the possibilities for synchronizing and integrating the hierarchy of *global and local distributed control algorithms*. An effective solution for traffic management requires coordinated systematic approaches that will: (1) effectively manage the large volumes of streaming data observations from heterogeneous sources, in conjunction with the existing historic values; (2) dynamically optimize the benefits of the data-awareness in the control algorithms; (3) balance the data-gathering process (observability) with the traffic control/management (controllability) for the purpose of optimizing the use of the sensing devices and networking resources.

In the rest of this extended abstract, in Section 2 we present an overview of the current state of the art in traffic management systems and point out a few observations regarding the deficiencies and possibilities of improvement. In Section 3, we present our vision about the potential role of the data mining in the efficient traffic management, and describe two specific topics that are focus of our current research which, we believe, constitute important components of any data-aware traffic control.

2. Background and Desiderata

A typical ATM system receives data from various sources, e.g.,: roadside vehicle identification sensors, inductive loop-detectors below the road surface, drivers who voluntarily provide location-in-time information via on-board equipment, vehicular ad-hoc networks, video detectors over roads, etc. [7]. Such data is used for: (1.) *traffic control* performed by adaptive traffic signals and phase timings [8,9]; and (2.) *traffic advisory*, ranging from roadside message signs and radio, through pre-trip and en-route guidance and navigation (e.g., Navteq [10], Roadnet [11]. A plethora of projects, prototypes and software tools exist (e.g., OPAC, RHODES, RTACLL, SIDRA [7]) that are currently used for controlling traffic lights at intersections in several cities in the US and worldwide. These generally aim at optimizing different traffic-related-metrics (e.g., average delay at signals, average trip-time, etc.). A generic diagram, illustrating the basic *predictive* paradigm for adaptive traffic management is illustrated in Figure 1, (cf. [7]). Under these settings the real-time information is used for prediction of the demands (e.g., duration of the green-light interval over a given horizon).

Another paradigm used in traffic management is the *reactive* one, where the real time detection of pre-defined events triggers the execution of corresponding actions that implement a particular optimization policy. A complementary approach for traffic management, also addressing the ever-increasing discrepancy between the traffic demands and the available road-infrastructure, is attempted by the Vehicle to Vehicle (V2V) and Vehicle to Infrastructure (V2I) paradigms [12, 13]. Intelligent Vehicles (IV) are viewed as a system that distributes the control over the vehicles themselves which, by sensing the environment and communicating the individual parameters with the traffic-peers, improve the overall operational efficiency in terms of fuel consumption, as well as CO and HC and emission [14]. Assistance can range from drivers advising, to automatically controlling the vehicle, i.e., its speed, distance from the other vehicles, etc. Several V2V/V2I control-frameworks with corresponding architectures have been proposed (e.g., PATH [15], Dolphin [16]), and actual projects have been initiated like, for example: – CarTALK, aiming at organizing the vehicles in an ad-hoc network for cooperative driver-assistance [17]; – SAFESPOT, aiming at improving safety and advanced detection of dangerous situations [18].

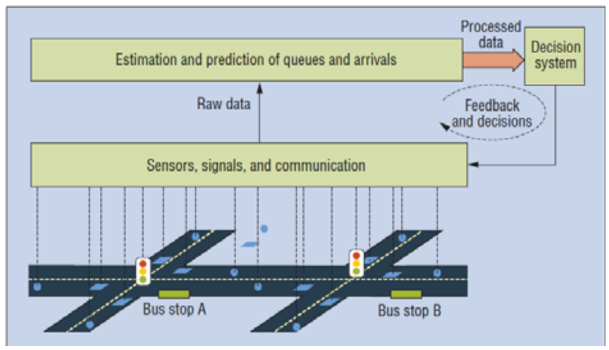

Figure 1. Real-time + Predictive Management

Current methodologies, however, have several limitations:

1. The decision and control activities used in most of the existing approaches that focus on regulating the traffic lights are *model-based*, and *do not fully exploit the automatic discovering and disseminating of knowledge and information* that can be obtained from the variety of *heterogeneous* streaming data sources.

2. From a complementary perspective – the V2V/V2I paradigms, offer advantages in terms of data-freshness and low communication latency, however, they still suffer from other drawbacks: the control-decisions based on the information received from the peers, have *limited spatio-temporal horizon of validity*.

3. Despite the wide variety of centralized and distributed algorithms for optimizing different parameters of relevance for efficient traffic management (such as road pricing or signal control [19]), there is a lack of *intelligent triggering mechanisms* that will enable balancing the trade-off between local vs. global control-decisions, in the context of the *information fusion*: maximizing the gain of the data gathering *(observability*) for the purpose of achieving a desired level of *controllability* within a minimal amount of time.

4. A variety of sensors are becoming available in automotive products for use of control, observation of individual vehicles, comfort and convenience. However, the impact of their deployment in larger scale transportation infrastructures (traffic lights, roadside signs, etc. [19]) has not been investigated thoroughly. Notwithstanding the recent advances in sensing technology and the information management in ad-hoc and wireless sensor networks, the spectrum of capabilities offered by networks of heterogeneous sensing devices has not been fully exploited for efficient traffic management.

3. Data Mining and Traffic Management

In order to fully capitalize on the recent advances in distributed data mining, miniaturization of computational and sensing devices, vehicular networks and wireless sensor networks, spatio-temporal data and events management in the domain of *efficient traffic management*, we postulate that *coordinated collaborative research efforts* are needed to:

(1) Develop new *theoretical foundations* that will: (1.a) maximize the quality of the distributed information fusion from heterogeneous devices. (1.b) provide novel computational paradigms that will synchronize the decision and control process with the data generation process for the purpose of optimizing the traffic parameters (e.g., flow).

(2) Develop novel *methods, tools and components* that will: (2.a) enable seamless integration of the distributed data mining and control algorithms with the sensing and communication devices, at different level of spatio-temporal granularity and across heterogeneous networks. (2.b) provide coordinated query optimization and data management, for heterogeneous streaming data. (2.c) provide unified approaches to hardware/ software co-design for novel programmable sensing and communication devices. Figure 2 presents a global view of the main components that, we believe, should be pursued in a collaborative manner, in order to effectively address the problem of the efficient and effective traffic management. The top part of the figure, adapted from [20], illustrates the desiderata of the VII Consortium, towards the goal of efficient traffic management. Positioning in this context, we seek to develop methodologies for context-aware management and mining of various spatio-temporal data, originating from heterogeneous sources and of different types, and use it to optimize the benefits of cooperative distributed control algorithms.

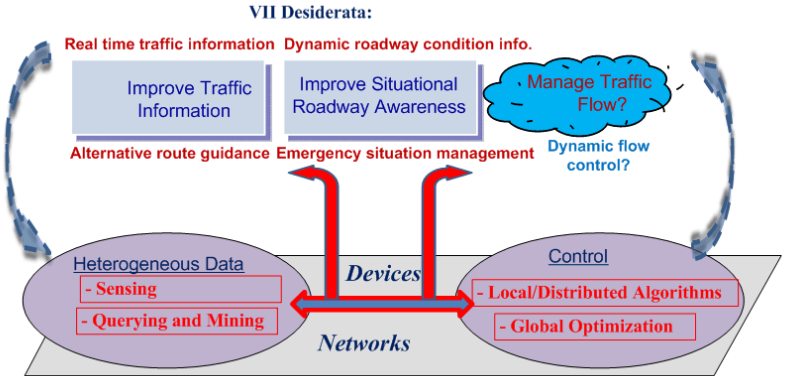

Figure 2. Data-Aware Traffic Control – System View

The *scalable and distributed data mining* has the potential to provide significant insights both in terms of formulating strategies and policies as well as for actionable patterns, rules and predictive models that can be used to adapt and enhance real-time traffic management in a manner that will couple the reactive and proactive behavior of the different system components. As an exemplary-scenario, consider a spatio-temporal analogy the well-known example of Amazon.com, where the associations amongst products and customers are successfully used in increasing recommendations. Along these lines, in our recent work [21], using the data from Gary-Chicago-Milwaukee corridor traffic sensors, we developed algorithms for associations discovery as well as predictive modeling in which certain correlated patterns are discovered across widely distributed geographical regions, as well as across different time periods. For example, as illustrated in Figure 3, traffic patterns at one region (O'Hare airport) may affect traffic patterns in 2 other regions (near Midway airport), 20-30 miles away and after 45-60min.

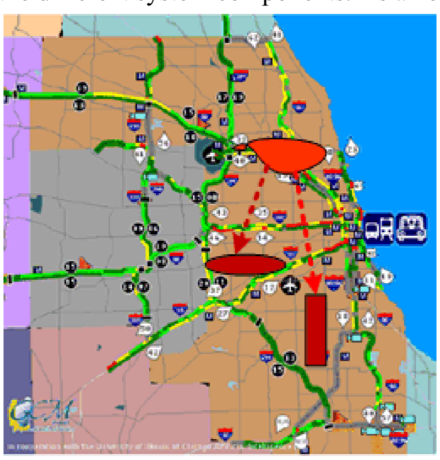

Figure 3. Correlating Traffic Patterns

However, these observations will not be of any actual use to the efficient traffic management, unless they are properly coupled with the distributed control algorithms. For example, it may be the case that slightly prolonging the red signal on the traffic lights on the intersections for the vehicles *moving towards* the Midway airport 20min. after detecting the congestion at O'Hare airport, will decrease the severity of the expected congestion within 2-3 miles from the Midway airport.

High-level patterns can be thought of as multidimensional clusters which provide a global view of the traffic patterns based on contributions of different context-dimensions such as *traffic volumes, time, space, external events* and so on. We believe that, as a first step, we need to divide the multidimensional data space in grids and use data mining techniques to perform clustering. The discovered clusters will provide patterns that can be used for policy and strategic decisions. For example, clusters that are formed based on certain events in Chicago, such as games, combined with a certain type of weather forecast, can be used for activating the controls for traffic management so as to minimize traffic densities in certain areas. On the other hand, if the weather forecast is different then the traffic patterns, with all other factors remaining the same, may form a different cluster resulting in a different decision. Although such patterns discovery will provide a high-level view, there may be different spatio-temporal dependencies between various patterns and clusters. In order to make decisions within each local area (local traffic grid, set of lights, or even within a vehicle), rules need to be developed for automated decision making.

We postulate that, towards the goal of detecting association rules of interest for efficient traffic management, one of the important tasks is identifying novel categories of spatio-temporal queries and devising efficient algorithms for their processing. As an example, the following query:

Q1: "Retrieve all the *platoons of vehicles* with the average speed between 35mph and 50mph that were *moving_towards* downtown on I-94, between 10:00AM and 11:30AM."

In [22] we presented efficient algorithms for processing variants of continuously moving *towards/away* and *along* predicates, however, **Q1** poses several novel challenges: (1) in addition to the motion-related predicate, now the system needs to detect *platoons;* (2) not all the platoons are of interest, but only the ones with specific speed-range, and along particular road-segments. Recent results on clustering trajectories data [23] and detecting of flocks of moving objects [24] do not properly address the aspect of the velocity of the objects, and we believe that approaches are needed that will efficiently detect the dynamics-based similarity of motions, allowing for rigid transformations (cf. [25]), potentially capitalizing on existing road-networks constraints [26].

When it comes to detecting correlations among spatio-temporal patterns, **Q1** itself can be useful as a "nested" component of other queries like, for example:

Q2: "For all the *regions* of radius > 3 miles and time-intervals > 30min in Chicagoland, for which the traffic flow was at least 40% below normal, retrieve all the platoons moving towards such regions, that had an average speed reduction of > 50% within at most 2 miles from the regions' boundaries."

The important observation is that the shape regions of interest for **Q2** may be subject to continuous deformations in their extents and, moreover, their number may vary over time. Clearly, a systematic exploration of different syntactic variants of novel categories of queries like, e.g., **Q1** and **Q2** are needed, along with the development of efficient algorithms for their processing, so that they can be applied for classifying traffic patterns, and detecting spatio-temporal association rules. We observe that, in addition to the need for evaluating such queries in a collaborative manner involving combined resources (e.g., distributed servers and the V2V networks), an important aspect is the detection of their *parameters*. Namely, based on the current traffic-data observations, it may be the case that the regions of radii > 2 (instead of > 3 in **Q2** above) miles may need to be considered.

Our envisioned traffic control system is based on the coordination of *actuation, sensing, computation (decision-making),* and *communication*. This will require a broad spectrum of design/implementation issues to be addressed collaboratively, ranging from high-performance computing systems (e.g., multicore architectures and reconfigurable devices), through P2P-like networks of traffic signals, to different (in-vehicle and roadside) sensors. However, at the heart of their efficient coordination is the ability to *react* to the detection of certain stimuli (events) and, pending the outcome of additional queries (*conditions*), execute *actions* in response. We believe that a paradigm is needed that will *not only react*, but will also *proactively orchestrate the subsequent tasks* and the usage of all the different resources. As an example, consider the following request:

Req1: "***When*** the average speed of platoons containing at least 10 vehicles, between County Line Rd. exit and Pulaski Rd. exit along I-55 inbound has declined by > 30%, ***if*** the average speed on the road-segments that have entry to I-55 has not decreased by > 10%, ***decrease by 10%*** the duration of the green-light interval on the nearest intersection to the ramp. Subsequently, ***when*** the density of vehicles on the streets within 2 miles along I-55 has increased by > 10%, ***if*** the average speed within 1 mile from the loop has dropped by > 40%, ***decrease by 15%*** the duration of the green-light interval on every intersection within 1 mile along I-55.

Req1 is an example of, what may be called an *evolving* trigger, i.e., a trigger forking children-triggers (cf. the term *Subsequently*). In [27], we presented the $(ECA)^2$ (Evolving and Context Aware Event Condition Action) paradigm for managing the reactive behavior in dynamic environments. In order to maximize the efficiency of the (impact of the) execution of such triggers, novel techniques for detecting association rules across different context dimensions are paramount. As an example, in the settings of **Req1**, the specifications of the event/condition parts and the action part need to be based on the actual observations of the historic spatio-temporal data in different geographic regions. The traffic control lights will need to organizing themselves in a P2P network with a *flexible spatio-temporal extent,* not only because of the specifications of **Req1**, but also for the purpose of balancing the *control-decisions* performed by, say, V2V flocks vs. the ones performed by the distributed (regional) servers, while optimizing the quality of the information. Such high-level control decisions coming from the central data servers will need a dynamic coupling with *decentralized* control algorithms, implemented on-board vehicles and on the traffic-signals. The benefits of such an approach are twofold: (1) local controls, e.g., on the traffic-signals, can react more quickly to disturbances or other sensed changes in the local traffic pattern (e.g., blocked lanes or a local increase/decrease of traffic flow in a *particular direction*); (2) signals equipped with such control algorithms can still accomplish intelligent signaling even when the central servers are down or otherwise unreachable. Because the decentralized controllers can react to sensed conditions before receiving answers from specific queries to the central data servers, they can operate on the faster time scales necessary for preventing significant performance loss in the presence of unexpected traffic disturbances.

References:

1. Sussman, J.M. *Intelligent Vehicle Highway Systems: Challenge for the Future*. in IEEE Micro. 1993.
2. Broucke, M. and Varaya., P., *The automated highway system: A transportation technology for the 21st century.* Control Engineering Practice, 1997.
3. *US_DOT_VII*. Available from: http://www.itsa.org/US_DOT_VII/c282/ITS_Resources/Library/US_DOT_VII.html.
4. Statistics, B.o.T. *Annual Wasted Fuel Per Person*. 2008; Available from: http://www.bts.gov/publications/national_transportation_statistics/html/table_04_28.html.
5. Ahn, K., et al., *Estimating Vehicle Fuel Consumption and Emissions based on Instantaneous Speed and Acceleration Levels.* Journal of Transportation Engineering, 2002.
6. *Traffic Safety Facts*. Available from: http://nhtsa.gov/staticfiles/DOT/NHTSA/NCSA/Content/RNotes/2008/811054.pdf.
7. Mirchandani, P. and Wang, F.-Y. *RHODES to Intelligent Transportation Systems*, in IEEE Intelligent Systems. 2005.
8. Lin, W.-C. and C. Wang. *An Enhanced 0-1 Mixed-Integer LP Formulation for Traffic Signal Control*. in IEEE Transactions on Intelligent Transportation Systems. 2004.
9. Mirchandani, P. and K. Head, *A Real-Time Traffic Signal Control System: Architecture, Algorithms, and Analyses.* Transportation Research Part C: Emerging technologies. Elsevier., 2005.
10. *NAVTEQ*. Available from: http://www.navteq.com/.
11. *Roadnet*. Available from: http://www.upslogisticstech.com/pub/products/Roadnet/.
12. Baskar, L.D., Schutter, B.D., and Hellendoorn, H., *Hierarchical Traffic Control and Management with Intelligent Vehicles*. in IEEE Intelligent Vehicles Symposium. 2007.
13. Bishop, R., *Intelligent Vehicles Technology and Trends*. 2005: Artech House.
14. Widodo, A., T. Hasegawa, and S. Tsugawa. *Vehicle Fuel Consumption and Emission Estimation in Envinonment-Adaptive Driving With of Without Inter-Vehicle Communications*. in IEEE Intelligent Vehicles Symposium. 2000.
15. Horowitz, R. and P. Varaiya, *Control Design of an Automated Highway System.* IEEE Special Issue on Hybrid Systems, 2000.
16. Tsugawa, S., et al. *An Architecture for Cooperative Driving of Automated Vehicles*. in IEEE Symposium on Intelligent Transportation. 2000.
17. Reichard, D., et al. *Cartalk2000 - safe and comfortable driving based upon inter-vehicle communication*. in IEEE Symposium on Intelligent Vehicles. 2002.
18. *SAFESPOT*. Available from: http://www.safespot-eu.org/.
19. Cipriani, E., and Fusco,G., *Combined signal setting design and traffic assignment problem.* European Journal on Operations Research, 155(3). 2003.
20. Robinson, R., *VII Strategies for Safety and Mobility*, in AASHTO RAC Annual Meeting 2006.
21. Liu, Y., et al. *A Scalable Distributed Stream Mining System for Highway Traffic Data*. in *10th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD)*. 2006.
22. Trajcevski, G., et al. *Dynamic topological predicates and notifications in moving objects databases*. in *Mobile Data Management*. 2005.
23. Lee, J.-G., et al. *TraClass: trajectory classification using hierarchical region-based and trajectory-based clustering.* VLDB. 2008
24. Gudmundson, J., and van Kreveld, M., *Computing longest duration flocks in trajectory data.* ACM-GIS. 2006..
25. Trajcevski, G. et al. *Dynamics-Aware Similarity of Moving Objects Trajectories* ACM-GIS. 2007.
26. Guting, R.H., et al. *Modeling and Querying Moving Objects in Networks. VLDB Journal*, 15(2). 2006.
27. Trajcevski, G., et al. *Evolving Triggers for Dynamic Environments*. in *Extending Database Technology*. 2006.